

# Text Analysis of Kurt Vonnegut with the HathiTrust & Voyant

Tassie Gniady  
David Kloster  
Guangchen Ruan  
Robert McDonald



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# Digital Arts and Humanities Workshop Series – Spring 2018

Fridays @ noon -- Scholars Commons IQ-Wall

Date	Topic	Presenter
Jan. 12	Intro to Digital Humanities	Tassie Gniady
Jan. 19	Intro to Visualization	Michael Boyles
Jan. 26	Intro to R	Tassie Gniady
Feb. 2	Augmented Reality	Chauncey Frend
Feb. 9	Text Analysis of Kurt Vonnegut w/ the HathiTrust & Voyant	Tassie Gniady & Robert McDonald
Feb. 16	Virtual Reality	Bill Sherman
Feb. 23	R for Twitter	Tassie Gniady
Mar. 2	Advanced Media	Chris Eller
Mar. 23	Network Graphs	David Kloster
Mar. 30	3D Object Acquisition & Printing	Jeff Rogers
Apr. 6	3D Photogrammetry	Tassie Gniady
Apr. 13	IQ-Table & Touch-Enabled Software Workflows	David Reagan
Apr. 20	Omeka S and 3D Collections	Tassie Gniady & Will Cowan

# Problem

Not all humanities researchers are comfortable with command line text manipulation & analysis, even when scripts have been prepared for them.

Jupyter Notebook for text cleaning



```
In [17]: import re
import string
import os
from nltk.corpus import stopwords

In [18]: #####
# Low level text cleaning routines
#####

'''
strip whitespaces for each line of a text string
'''
def stripLines(text):

    lines = text.splitlines()

    # strip each line
    lines = [x.strip() for x in lines]

    # eliminate empty lines
    lines = [x for x in lines if x]

    return '\n'.join(lines)

'''
Convert a text string to lower case
'''
def toLower(text):

    return text.lower()

'''
Remove digits from text
'''
def removeDigits(text):

    p = re.compile('\d+')
    text = p.sub('', text)

    return text

'''
Remove punctuations from a text string
'''
def removePunc(text):

    translator = str.maketrans('', '', string.punctuation)
```



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# Solution

Make a text analysis tool with a good GUI available in the capsule. Enter [Voyant](#).

## Digital humanities

From Wikipedia, the free encyclopedia

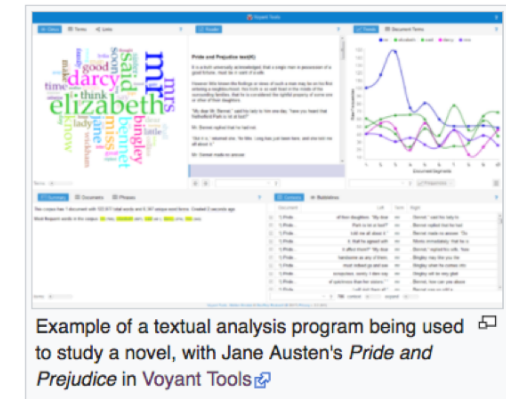
**Digital humanities (DH)** is an area of scholarly activity at the intersection of [computing](#) or [digital technologies](#) and the disciplines of the [humanities](#). It includes the systematic use of digital resources in the [humanities](#), as well as the reflection on their application.<sup>[1][2]</sup> DH can be defined as new ways of doing scholarship that involve collaborative, transdisciplinary, and computationally engaged research, teaching, and publishing.<sup>[3]</sup> It brings digital tools and methods to the study of the humanities with the recognition that the printed word is no longer the main medium for knowledge production and distribution.<sup>[3]</sup>

By producing and using new applications and techniques, DH makes new kinds of teaching and research possible, while at the same time studying and critiquing how these impact cultural heritage and digital culture.<sup>[2]</sup> Thus, a distinctive feature of DH is its cultivation of a two-way relationship between the humanities and the digital: the field both employs technology in the pursuit of humanities, research, and subjects technology to humanistic questioning and interrogation, often simultaneously.

### Contents [\[hide\]](#)

- [1 Definition](#)
- [2 History](#)
- [3 Values and methods](#)
- [4 Tools](#)

As the Wikipedia page about digital humanities has evolved, Voyant has taken top billing as a popular tool.



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



## Solution (cont)

It has always been possible to download the source code for Voyant and load it into an HTRC data capsule, compile it, and then run the tool. However, there are some java settings to be tweaked, and the user needed to have set up a capsule with enough memory to run the application. To truly lower the barrier to entry, we asked the HTRC to load Voyant into every data capsule, and put a link in the sidebar.



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services

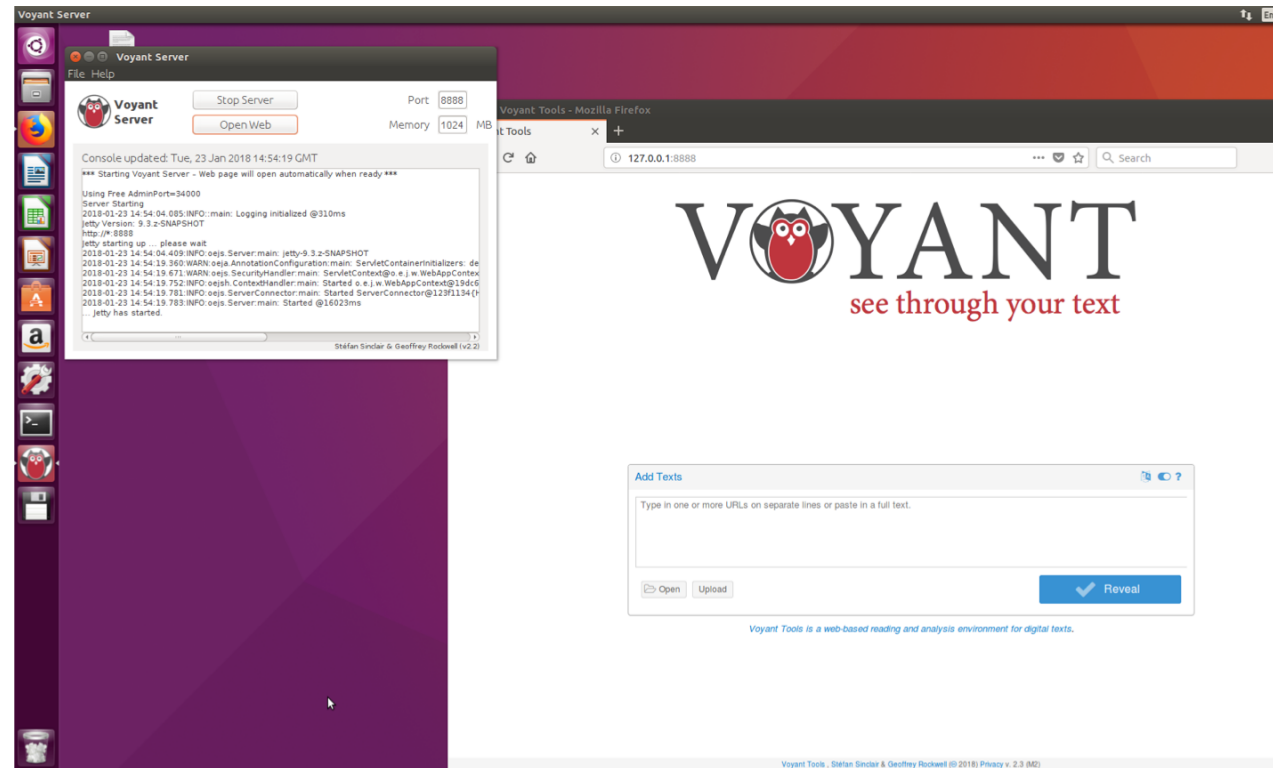


**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# Double-click!



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

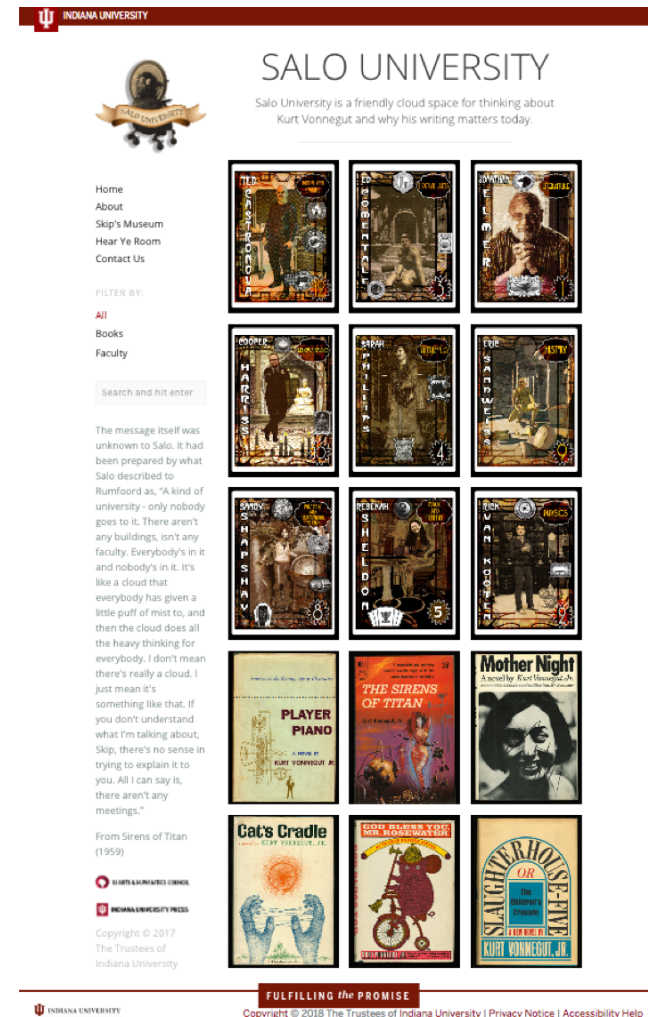
INDIANA UNIVERSITY



# Use Case

Salo University, an Open Humanities project at Indiana University brings together nine scholars to read and post their thoughts on six of Vonnegut's novels. In keeping with Vonnegut's far-reaching scientific imagination, his texts are prime fodder for computational analysis.

Using *Cat's Cradle* as a sample text, we upload it into Voyant.



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



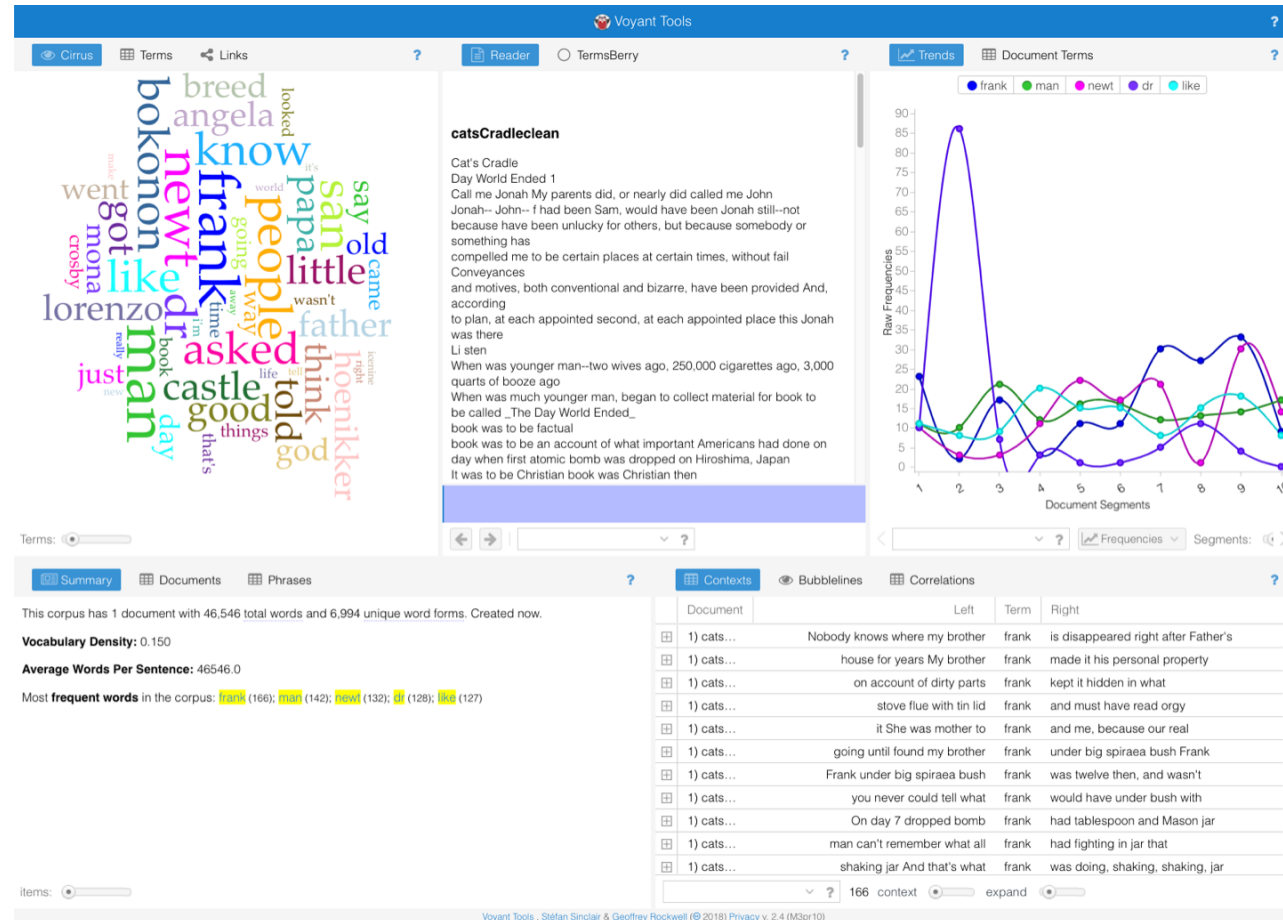
PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY





# First Look



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services

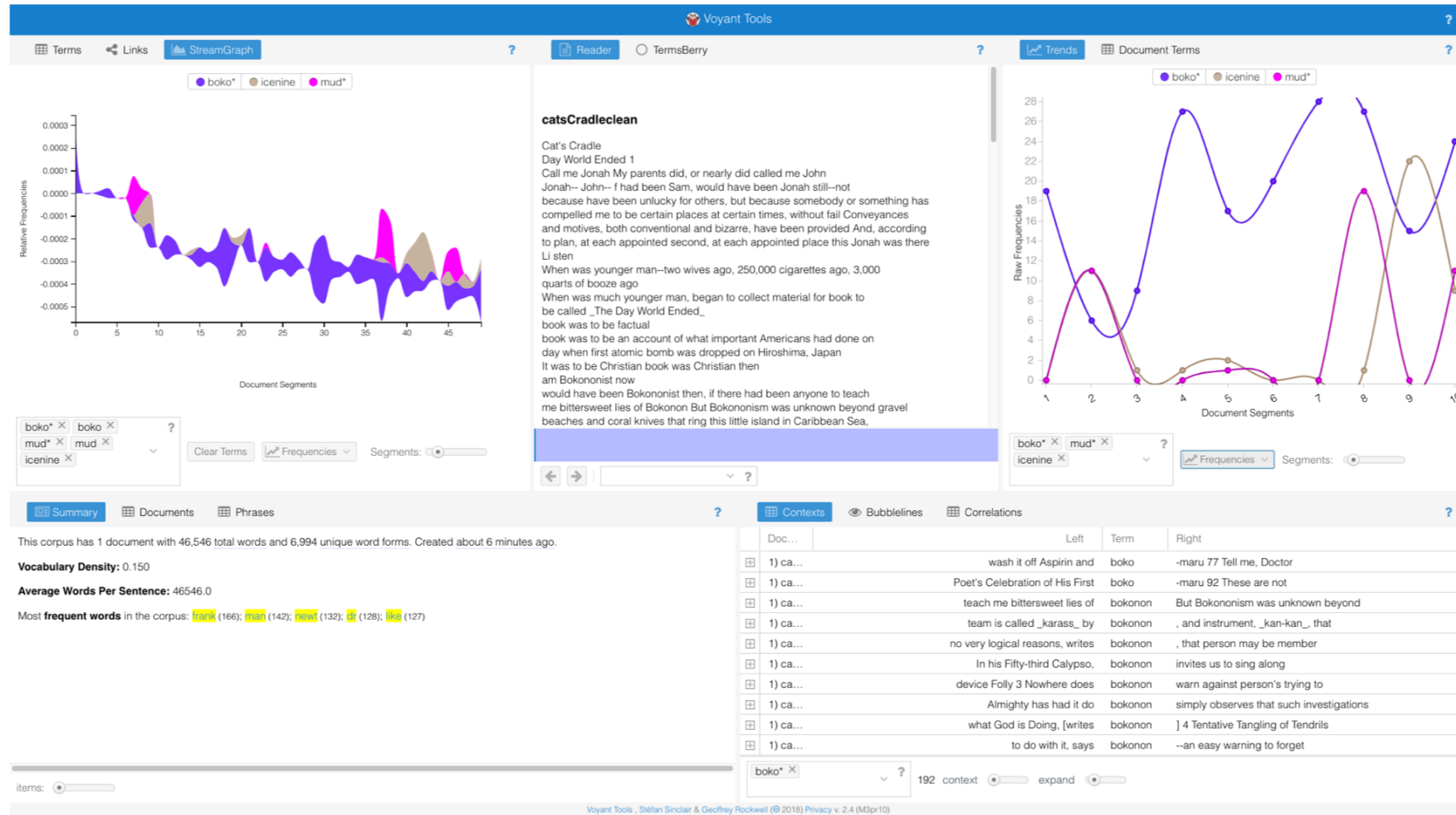


PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# Another Tailored View



RESEARCH  
TECHNOLOGIES

INDIANA UNIVERSITY  
University Information Technology Services



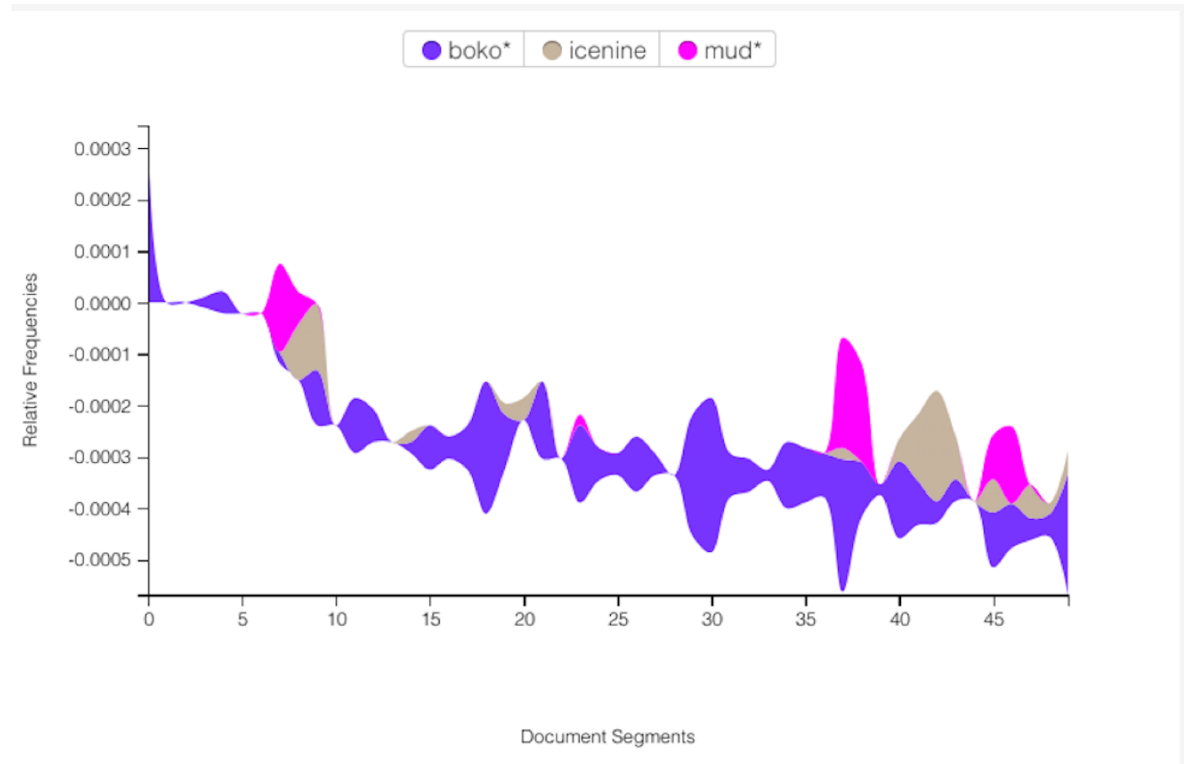
PERVASIVE TECHNOLOGY  
INSTITUTE

INDIANA UNIVERSITY



# Stream Graph Explicated

While this novel focuses on *both* technology and religion, Bokononism dominates Vonnegut's syntax in a way that technology does not. It winds its way through the novel, present even when the science of ice-nine is being discussed. Mud may be seen as a bridge between the two as it relates the messiness of war to the messiness of creation. Thus, four of the peaks related to mud coincide with ice-nine, revealing that the unstructured nature of dirt + water is always in discourse with the structure of frozen H<sub>2</sub>O.



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services

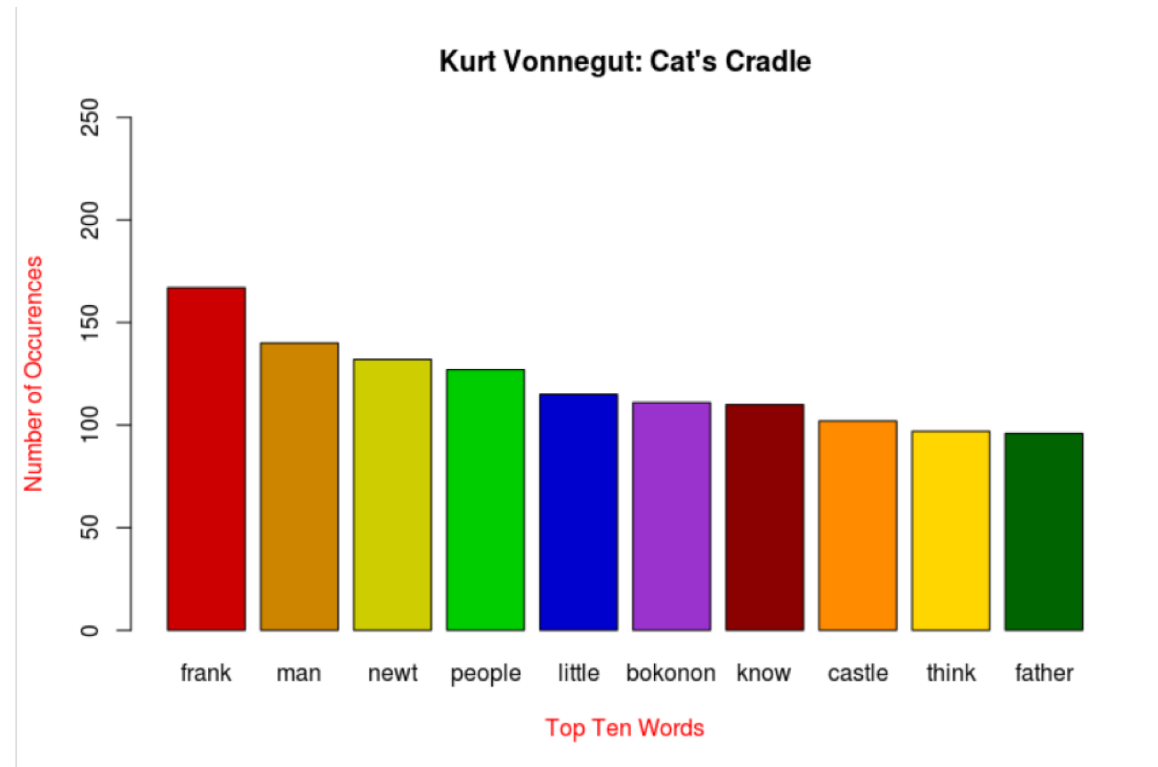


**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# Cat's Cradle



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services

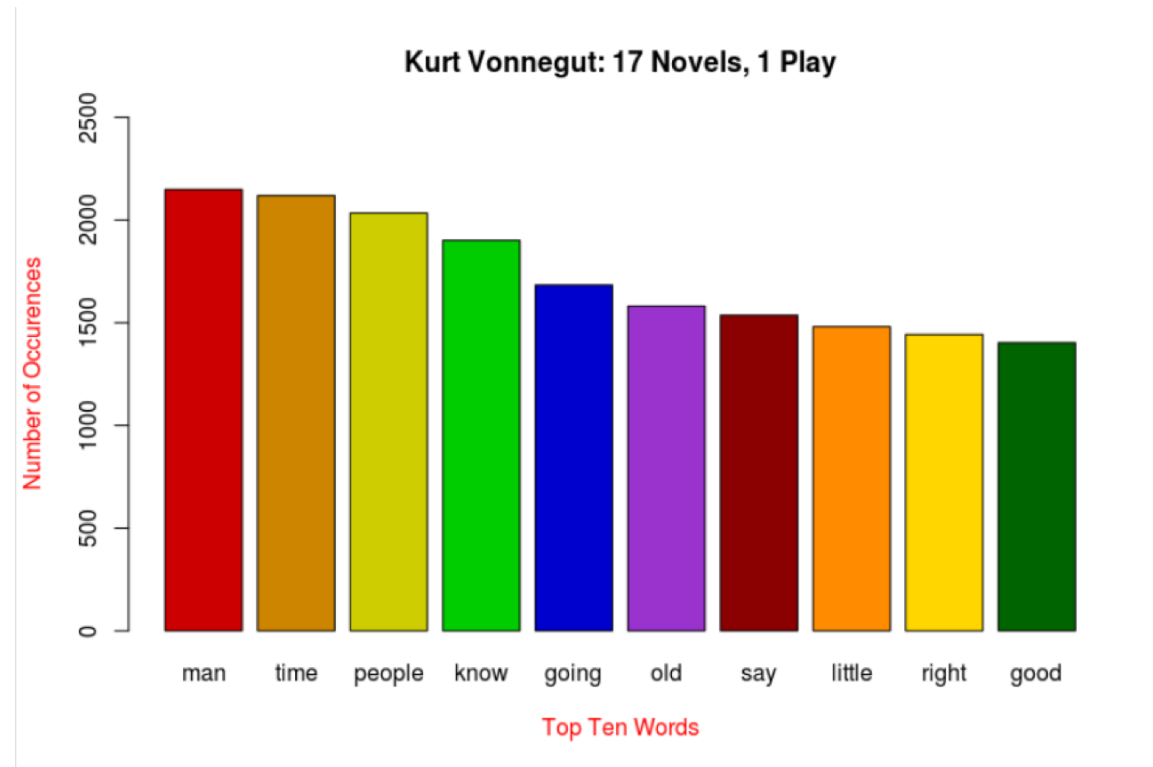


**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



# HathiTrust Vonnegut Corpus



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



## Next Steps

- The six novels read by the group will be topic modeled individually and together (another tool available on Voyant) and presented to the scholars for feedback
- The entire corpus of Vonnegut will be topic modeled and analyzed by one partner from the group of Vonnegut experts
- The corpus of Ray Bradbury, another Indiana-born science fiction author, will be topic modeled and compared to Vonnegut
- The results of these analyses will be presented at a [Granfalloon: A Kurt Vonnegut Gathering](#) in May as a demonstration to experts of the possibilities of non-consumptive computing on in-copyright texts with low barrier to entry



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY



Thanks!

If you have questions about this work, please feel free to contact [cyberdh@iu.edu](mailto:cyberdh@iu.edu)

Canvas Site for Workshop Series: <http://go.iu.edu/1SKs>



**RESEARCH  
TECHNOLOGIES**

INDIANA UNIVERSITY  
University Information Technology Services



**PERVASIVE TECHNOLOGY  
INSTITUTE**

INDIANA UNIVERSITY

